

TDT4173: Machine Learning and Case-Based Reasoning

Exercise 1

Published: August 28th, 2012

Delivery deadline: September 6th, 2012

I

1. Give two examples of relevant machine learning problems and describe them as “well-posed learning problems”.
2. What is inductive bias (for a learning method)? Why is it so important in machine learning? The candidate elimination algorithm for learning in version spaces and learning of decision trees with ID3 are two different learning methods. What can you say about the inductive bias for each of them?

II

You are supposed to design a program for learning the estimate of the *V function* for the tic-tac-toe game (like in exercise 1.5 in Tom Mitchell’s book). Implementing this program in code is not required, but pseudocode can be useful to illustrate some of the concepts.

1. What would be a good target function representation for learning to play tic-tac-toe?
Hint: $\hat{V}(x) = w_0 + w_1x_1 + \dots + w_nx_n$.
2. How would you represent the tic-tac-toe board in a programming language of your choice?
3. How would you detect the final win, loss or draw situations?
4. How would you calculate the features (x_i) you chose for your representation?
5. How would you determine which move to play next for a given board position?
6. How would you use training examples to improve your target function?
7. What are the main differences between using this approach for tic-tac-toe game playing and using search approaches known from AI (e.g. minimax, alpha-beta pruning)?

III

Candidate elimination (CE) is a learning method that tries to generate a description which is consistent with all positive — and no negative — examples in the training set. In this exercise, the following attributes and possible values to describe animals are given.

Attribute	Values
Body	Hair, Scales, Feathers
Birth	Live, Egg
Eats-Meat	True, False
Flies	True, False
Teeth	Pointed, Flat, None

The task is to learn the class *Mammal* from the following training examples.

Example	Body	Birth	Eats-Meat	Flies	Teeth	Mammal
1	Hair	Live	False	False	Flat	True
2	Feathers	Egg	True	True	None	False
3	Hair	Live	False	False	Pointed	True

1. Use the CE algorithm for version space learning and show S and G after each new training example. After learning on the training examples, use the system to classify the following test examples.

Example	Body	Birth	Eats-Meat	Flies	Teeth	Mammal
1	Hair	Live	False	False	None	??
2	Feathers	Egg	False	True	Pointed	??
3	Scales	Egg	True	False	Flat	??

2. What can the system say about the classification of the three new examples? Explain why.
3. Assume that the system can ask for another training example. Which criteria should the system use to choose the training example? Give an example combination of attribute values that satisfies this criteria.

IV

Another method that can be used to identify concepts is using sets of IF-THEN rules. One way to create such rule sets is to use a sequential covering algorithm to incrementally generate new rules until the target concept is adequately described.

1. What rules would you create to gradually cover the training examples given for the *Mammal* domain in this way? Why did you pick that set?
Hint: For this task you can manually choose which rules to add without using a specific rule learning algorithm.
2. Assume that three more training examples $\{Scales, Egg, False, False, Pointed\} = True$, $\{Scales, Live, True, False, Flat\} = True$ and $\{Hair, Live, True, True, Pointed\} = True$ are added for the *Mammal* domain.

How does this affect what rules you would pick to describe the *Mammal* concept? Explain how you gradually add new rule(s) to expand the existing set. How does this differ from using the CE algorithm?